DOCUMENT RESUME

ED 384 667 TM 023 974

AUTHOR Bridgeman, Brent; Lewis, Charles
TITLE Sex Differences in the Relationship of Advanced Placement Essay and Multiple-Choice Scores to Grades in College Courses.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-91-48
PUB DATE Aug 91
NOTE 55p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Advanced Placement; Biology; *College Freshmen; English; *Essay Tests; European History; Females; *Grade Point Average; Higher Education; Males; *Multiple Choice Tests; Prediction; Scores; *Sex Differences; Student Placement; Test Results; *Test Use; Test Validity; United States History
IDENTIFIERS *Advanced Placement Examinations (CEEB)

ABSTRACT

Essay and multiple-choice scores from Advanced Placement (AP) examinations in American History, European History, English Language and Composition, and Biology were matched with freshman grades in a sample of 32 colleges. Multiple-choice scores from the American History and Biology examinations were superior to essays for predicting overall grade point average, but essay scores were essentially equivalent to multiple-choice scores for predicting grades in history, English, and biology. In history courses, males and females received comparable grades and had nearly equal scores on the AP essays, but the multiple-choice scores of males were about half of a standard deviation higher than the scores of females. To the extent that the AP history examinations are intended to emulate performance in college history courses, placing greater weight on the essay component of the AP history examinations would reduce sex differences without compromising content or predictive validity. Eleven tables present study findings. Appendix A lists score conversions, and Appendix B presents an additional three tables of correlations of AP scores and grades. (Contains 14 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

RR-91-48

RESEARCH REPORT

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H. I. BRAUN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SEX DIFFERENCES IN THE RELATIONSHIP OF ADVANCED PLACEMENT ESSAY AND MULTIPLE-CHOICE SCORES TO GRADES IN COLLEGE COURSES

Brent Bridgeman
Charles Lewis

BEST COPY AVAILABLE

ETS

Educational Testing Service
Princeton, New Jersey
August 1991

ED 384 667

TM023974

Sex Differences in the Relationship of Advanced Placement Essay and Multiple-Choice Scores to Grades in College Courses

Brent Bridgeman

Charles Lewis

Educational Testing Service

Copyright © 1991. Educational Testing Service. All rights reserved.

Acknowledgements

Thanks to Leonard Ramist for assembling the database and supporting access to it. Programming and data analysis were ably performed by Annette Turner.

Abstract

Essay and multiple-choice scores from Advanced Placement (AP) examinations in American History, European History, English Language and Composition, and Biology were matched with freshman grades in a sample of 32 colleges. Multiple-choice scores from the American History and Biology examinations were superior to essays for predicting overall GPA, but essay scores were essentially equivalent to multiple-choice scores for predicting grades in history, English, and biology. In history courses, males and females received comparable grades and had nearly equal scores on the AP essays, but the multiple-choice scores of males were about half of a standard deviation higher than the scores of females. To the extent that the AP history examinations are intended to emulate performance in college history courses, placing greater weight on the essay component of the AP history examinations would reduce sex differences without compromising content or predictive validity.

Sex Differences in the Relationship of Advanced Placement Essay and Multiple-choice Scores to Grades in College Courses

Essay examinations assess productive and organizational skills that can not be measured with multiple-choice questions, but they require time consuming and expensive scoring sessions that can be run only with trained experts in the subject area of the examination. On the other hand, multiple-choice tests are easy to score with machines. The two types of tests also differ in their coverage of the content domain. Essay examinations usually require an in-depth understanding of a few content areas while multiple-choice examinations survey a broader range of topics. Because of measurement error created by subjective scoring and by the relatively narrow coverage of the content domain, essay tests are usually substantially less reliable than multiple-choice tests in the same general subject area.

Previous research comparing essay and multiple-choice tests as measures of writing ability suggests that the two kinds of assessment are largely overlapping, but that each also assesses some unique features (Breland & Gaynor, 1979; Quellmalz, Capell, & Chou, 1982; Breland, Camp, Jones, Morris, & Rock, 1987). However, much less is known about the relative contributions of the two types of assessment techniques for prediction of subject matter mastery as opposed to assessment of writing ability.

One of the important, and currently unexplained, differences between multiple-choice and essay assessments of subject matter mastery is that, in several different subject areas, essay assessments produce smaller sex differences than do multiple-choice tests. Evidence from Great Britian (Murphy, 1982), Australia (Bell and Hay, 1987), and Ireland (Bolger and

Kellaghan, 1990) consistently indicates that males have a relative advantage on multiple-choice tests. Mazzeo, Schmitt, and Bleistein (1989) found a similar male advantage on several of the Advanced Placement (AP) examinations that are taken by high school students who are seeking college credit or placement into advanced college courses based on college-level courses that they have completed. In several different subject areas, males and females had nearly equal scores on the essay portion of the examination while males had significantly higher scores on the multiple-choice portion. This difference remained even after correcting for the differential reliability of the two question types, and removing items from the multiple-choice test on which males did particularly well (i.e., high DIF items) had very little impact on the observed sex differences. Differences were especially striking on the United States History examination, with estimated true score means for males and females essentially equivalent on the essays (difference of less than .02 in standard deviation units) but with the mean for males more than .3 standard deviation units higher than the mean for females on the multiple-choice portion of the test. This difference is particularly important because American History is one of the largest AP programs, testing over 50,000 students a year. Similarly large differences were found on the European History examination. Smaller differences were found on the Biology examination (standardized difference on the essay section was .16 while on the multiple-choice section the difference was .33) and on the English Language and Composition examination (difference of .02 on the essay section and .17 on the multiple-choice section).

In a study of the validity of the Advanced Placement Examination in Biology for predicting grades in college biology courses, Bridgeman (1989)

found that the essay and multiple-choice scores predicted grades about equally well for males, but that for females predictions were superior using the multiple-choice scores. However, sample sizes were small and there was considerable variation among colleges.

One purpose of the current study was to assess the validity of the essay and multiple-choice sections of selected AP examinations for predicting success of males and females in college courses. It should be noted that the primary purpose of the AP examinations is to certify the acquisition of skills and knowledge taught in specific college-level courses taken in high school, not to predict success in those or similar courses taken in college. Thus, content not predictive validity is of paramount importance. Nevertheless, a reasonable part of the construct validity of such examinations is that success in the components of the examination should be correlated with success in related courses taken in college. If there were a major discrepancy between predictions from the essay or multiple-choice components of an examination (or between predictions for men and women on either component) it would not automatically mean that the offending component should be dropped, but it would suggest that it be closely scrutinized. A second purpose of the study was to determine whether the sex differences observed on the AP multiple-choice scores (and hence composite scores) were reflected in similar sex differences in college grades.

The AP examinations selected for study were those with a relatively large discrepancy between the multiple-choice sex difference and the essay sex difference. An additional criterion was that the test be taken by enough students to have some hope of finding sufficient numbers of students for analysis when looking at particular courses within a college. Thus, for

example, the Physics B examination, which has a much larger sex difference on the multiple-choice section than on the free-response section, was rejected because too few students (especially female students) took it. The Calculus AB and English Literature examinations, although taken by a large number of males and females, were rejected because the sex differences on the multiple-choice and free-response sections were virtually identical. The examinations selected were American History, European History, English Language and Composition, and Biology.

METHOD

Sample

Freshman grades from 45 colleges were obtained from a database that had been assembled to support a variety of validity studies. Colleges in which fewer than five male and five female students had taken one of the target AP examinations in combination with the appropriate freshman courses were eliminated from the sample, resulting in a final sample of 32 colleges. The colleges in the database include both public and private institutions that use the Scholastic Aptitude Test (SAT) as part of the admissions process. The grades were from students who entered college in the fall of 1985. The database contains grades in individual courses as well as summary averages that group courses in related fields. For example, a history average represents the grade point average of a student in all of the history courses that the student took during the freshman year. In some cases, courses intended for majors that included labs were separated from more general courses, and courses that are generally intended for more advanced students were separated from regular freshman courses. Grades were coded on a 13 point

scale (F=0, D-=.7, D=1, D+=1.3, C-=1.7, ...A=4, A+=4.3). In addition to grades, the database included SAT scores and self-reported high school grade point average (HSGPA).

AP scores from the selected examinations were added to the database and matched by social security number. AP tests from 1984 and 1985 were included. Although most AP examinations are taken at the end of the senior year in high school (i.e., 1985 for students who began college in 1985), a notable exception is the American History examination which is typically taken at the end of the junior year because most students take American history as an 11th grade course. Of the 53,859 students in the original database of college grades, AP scores were located for 7626 students (about 14%); of these 7626 students, 6243 took one AP examination, 1237 took two AP examinations, and the remaining 146 took three or more AP examinations.

AP test descriptions

United States History. The multiple-choice portion of this examination consisted of 100 items with five answer choices for each item. Examinees were allowed 75 minutes to answer these questions. This section was formula scored (score=number right-¼ number wrong) with negative formula scores converted to 0. The essay portion of the test consisted of two essays. For the first essay, examinees were provided with a set of documents and asked to construct an argument based on the documents. In order to receive an above average score the candidate had to make reference to historical facts that were not directly discussed in the documents provided. For the second essay, examinees were asked to choose one of six thematic history questions that were presented.

Each of the two essays was scored by a different reader using a 0-15

scale; thus, essay scores could range from 0 to 30. The composite score was created by multiplying the multiple-choice formula score by .9, multiplying the essay score by 3 and summing the two weighted scores. Thus, the two sections received nominally equal weights in forming the composite score (each contributed a maximum of 90 points), but because the standard deviation of the multiple-choice section was slightly larger (14.8 vs. 12.7 in 1984) the multiple-choice section actually had a slightly greater importance in the determination of relative rankings on the composite score. The chief reader and ETS professional staff then transformed the composite score to the 1-5 scale that was reported to colleges. This transformation was based to a large extent on an equating of the multiple-choice scores on a given form with the multiple-choice scores on an earlier form through a set of items common to both forms. Appendix A contains a table for converting composite scores (used in this report) to the 1-5 scores that are reported to colleges.

Reliability of the multiple-choice scores, as estimated by KR-20, was .90 in 1984 (Eignor, Flesher, and McClean, 1984) and .89 in 1985 (Livingston, McClean, and Flesher, 1985). The coefficient alpha reliability of the essay scores was .54 in both years; because each essay is read by a different reader, this estimate includes both differences among readers and differences among topics as sources of unreliability. Reader reliability alone was about .79. Correlation between the multiple-choice and essay sections was .48 in 1984 and .53 in 1985.

European History. The general format and scoring rules for this examination were nearly identical to the American History examination except that candidates were not expected to use outside knowledge in answering the document-based essay question. The KR-20 reliability of the multiple-choice

score was .91 and the coefficient alpha reliability of the essay score was .44. The correlation between the two sections was .50 (Mazzeo and Flescher, 1985a).

English Language and Composition. This examination consisted of 65 five-option multiple-choice items that were formula scored in the manner described above. This portion of the exam took 75 minutes and contained two types of questions. One type tested a student's ability to manipulate syntax by recasting sentences, and the other type contained questions that asked the student to analyze the rhetoric, style, and content of prose passages. For the essay portion of the examination, the examinee was given a 15-minute reading period, and then asked to answer three questions in 90 minutes. Each question required a response in a different rhetorical mode. Each essay was graded by a different reader using a 9-point scale, so raw scores ranged from 0 to 27. The composite score was created so that the multiple-choice section contributed 40 percent and the essay section contributed 60 percent of the maximum total score of 150 (multiply the multiple-choice score by .923, multiply the essay score by 3.333, and sum). Reliabilities of the multiple-choice and essay sections were .88 and .56 respectively. The correlation between the two sections was .51 (Livingston, Karatka, and Bleistein, 1985).

Biology. In 1985, the 90-minute multiple-choice portion of this examination consisted of 120 five-option items that were formula scored. Three topics were assessed with 40 items on each topic: (A) Cellular and Molecular, (B) Organismal, and (C) Populational. On the 75-minute essay section there were three pairs of questions, one pair from each of the above topics. The candidate was instructed to choose one question from each pair. Each of the three essays was graded on a 15-point scale. Multiple-choice

scores were multiplied by .625 and essay scores were multiplied by 1.667 so that the two portions of the examination made nominally equal contributions to the total score of 150. Reliability of the multiple-choice section was .93 while the coefficient alpha reliability of the essay section was .66 (Mazzeo and Flesher, 1985b). Reader reliability alone was about .85. The correlation of essay and multiple-choice scores was .73.

RESULTS AND DISCUSSION

Correlation of AP examinations with college GPAs

For each college that contained at least five male and five female students who took the relevant AP examination, a correlation was computed between the overall college grade point average (GPA) and the essay and multiple-choice AP scores. These within-college correlations were transformed to z scores with Fischer's r to z transformation and were weighted by n-3. The weighted mean of the z scores was then transformed back to r. The correlations are presented in Table 1. Except as noted for the 1984 American

---

Insert Table 1 about here

History examination, the tests were administered in 1985 and reflected performance at the end of the senior year in high school. For a particular examination, the correlations for the multiple-choice section and essay section were generated by the same people and may be readily compared to determine whether one question type is a significantly better predictor of overall GPA (Dunn & Clark, 1969). As indicated in the table, for Biology and for both years of the American History test the multiple-choice scores are significantly better predictors of GPA. Furthermore, for all practical

purposes the composite score for these examinations was no better than the multiple-choice score by itself. Patterns within sex essentially replicated the combined sex results although they sometimes fell short of statistical significance in the smaller groups. For the European History and English Language examinations the differences between predictions from multiple-choice and essay questions were not significant, and the composite appeared to be better than either section by itself.

Across examination comparisons must be made more cautiously because people are not randomly assigned to take particular examinations, and the characteristics of students who choose to take the biology examination, for example, might be quite different from the students who choose to take American History. Given this caveat, there is a surprising degree of similarity among the examinations as predictors of GPA; across examinations the correlation of the composite score with GPA ranged from .31 to .36.

Correlations of the AP examinations with grade point average only in social science courses are presented in Table 2. The social science grade

---

Insert Table 2 about here

---

average includes such courses as anthropology, psychology, and sociology taken during the freshman year, but does not include courses in any of the target disciplines (history, English, and biology). The pattern of correlations for this set of courses was nearly identical to the pattern for the full freshman grade point average.

Correlation of AP history examinations with college history grades

Within each college that contained at least five male and five female students who took the 1984 AP American History examination and at least one

history course, a correlation was computed between the history grade point average and the essay and multiple-choice AP scores. These within-college correlations were averaged in the same manner described above. Similar procedures were followed for students who took the 1985 American History or European History examinations.

Insert Table 3 about here

The averaged correlations are presented in Table 3; Appendix B presents the correlations separately for each college in the sample. For prediction of history grades, the AP American History essays appear to be at least as good as the multiple-choice questions. Although it is difficult to prove the null hypothesis, it should be noted that the standard error of the difference between the '84 American History essay and multiple-choice correlations was only .036 with the 95% confidence interval ranging from -.07 to .07. Results for the European History examination were less clear because of the relatively small sample size. The apparent advantage for the multiple-choice items on this examination was not statistically significant.

Correcting the essay score for unreliability (recall that the essay score is much less reliable than the multiple-choice score) presents an even stronger case for the potential value of an essay score. In the large '84 American History sample, the estimated correlation of a perfectly reliable essay test with history grades (which were not corrected for unreliability) was .44 while the correlation of a perfectly reliable multiple-choice test

with grades was .31[1]. Assuming that the standard error of the corrected scores was reasonably similar to the standard error of the uncorrected scores, this would represent a significant advantage of a perfectly reliable essay test over a perfectly reliable multiple-choice test for predicting history grades.

Although sex differences in prediction for '84 American history and '85 European history appeared to be trivial, '85 American history appeared to be anomalous with multiple-choice questions a significantly better predictor for women; the sex difference for the essay showed a nonsignificant trend in the opposite direction. No explanation of this difference was uncovered. There were no changes in the test specifications from 1984 to 1985. Although the essay topics were different in the two years, the discrepancy was noted for both the document essay and the choice essay in 1985. The document essay correlated .24 with history grades for males and .01 for females; the choice essay correlated .29 for males and .06 for females. For the '84 examination, there were no sex differences in prediction for the document essay (.28 vs .23) or the choice essay (.22 vs .21). Differences in overall group ability or homogeneity across the two years would not lead to the kind of sex by

---

[1]In order to make the correction, within-college reliability estimates for both multiple-choice and essay scores were needed. The within-college reliability of the essay could be directly estimated with KR-20 because the correlation between the two essay scores (document and choice) within each college could be computed from the available scores and averaged over colleges. However, the reliability of the multiple-choice scores could not be directly determined because within-college item level data were not available. But reliability estimates for both the essay and multiple-choice scores were available from the national sample. The within-college reliability for the multiple-choice score was estimated by assuming that the ratio of true score variance in the college sample to true score variance in the national sample is the same for both essay and multiple-choice tests. Because this ratio could be computed for the essay score, and the reliability of the multiple-choice score in the national sample was known, the reliability of the within-college multiple-choice score could be estimated.

question type (multiple-choice vs essay) interaction observed; in any event, means and standard deviations were comparable in the two years. Because the '85 American history sample is relatively small, it may be unwise to pay too much attention to an apparent sex interaction that failed to replicate in a sample that was more than twice as large.

The relative advantage of essays for predicting history grades, as contrasted with predicting overall GPA, may relate to the likelihood that history courses will be graded with essay tests. A telephone survey of the largest colleges in the sample confirmed that essay tests were always the primary grading criterion in the history courses while other courses frequently relied on multiple-choice examinations. The importance of the similarity of assessment methods for the predictor and criterion is underscored by a comparison of the AP scores with high school grade point average and SAT verbal (SAT-V) and mathematics (SAT-M) scores. High school grade point average is based to some extent on performance on multiple-choice classroom tests but essay tests and other assessment methods are also included; the SAT is strictly a multiple-choice test. Table 4 shows that for

---

Insert Table 4 about here

---

the two largest samples ('84 American history and '85 biology) essay and multiple-choice AP scores were about equally correlated with high school grade point average (HSGPA), but AP multiple-choice scores were substantially more related to SAT scores than were AP essay scores.

Correlation of English Language AP scores with college English grades

The correlations of the scores from the AP English Language and Composition examination with college English grades are presented in Table 5.

The results were very similar to the history findings; despite its lower

---

Insert Table 5 about here

---

reliability, the essay score was at least as good a predictor as the multiple-choice score in both sex groups, and the composite score was a slightly better predictor than either of the individual scores. The reasoning that explained the history results also applies here; when the criterion score (i.e., English grade) is largely determined by the quality of student writing then the score on an essay examination is a relatively good predictor.

Correlation of AP Biology with college biology grades

For this analysis only regular courses in the biological sciences that were normally open to freshman students were considered; courses that were specifically designated for students majoring in biology were excluded. As noted previously (Bridgeman, 1989), very few students use their AP experience to enroll in advanced biology courses during their freshman year. In the current set of colleges, a sample of only 43 students in advanced courses was identified. This sample was deemed to be too small for meaningful analysis. Also consistent with the previous study was the finding that students are often encouraged not to take biology courses before their sophomore year. Thus, for example, at one of the larger colleges in the sample 62 students were identified who had taken the AP biology examination, but none of these students was enrolled in a regular biology course.

As indicated in Table 6, correlations were comparable for the essay and

---

Insert Table 6 about here

---

multiple-choice scores, and sex differences were not significant. The absence

of a difference between the correlation from the essay test and the correlation from the multiple-choice test in the female sample failed to replicate the finding of such a difference in an earlier study (Bridgeman, 1989). Combining the data on female students from both studies into a single analysis (n=370) resulted in a correlation with biology grades of .34 for the essay score and .45 for the multiple-choice score with a standard error of the difference of .06, thus just reaching the conventional standard of statistical significance (t=2.19, p<.05). But the finding is not very robust; removing the single course with the largest difference in the first study (which was also the largest course [n=51]) from the combined sample resulted in correlations that did not differ significantly (rs of .38 and .42 for the essay and multiple-choice scores respectively). As noted in the earlier study, there is substantial variation among colleges in these patterns of correlations; in 13 of the independent samples of female students in the combined studies the essay score was a better predictor than the multiple-choice score while the opposite pattern was observed in the other 18 samples.

Summary of correlational results

When grades in history, English, or biology courses are the criterion, the essay scores from the corresponding AP examinations predict about as well as the multiple-choice scores despite their lower reliabilities. Sex differences in correlational patterns appeared to be trivial, especially in the large English and '84 American history samples. For unknown reasons, the essay was a poor predictor for females in the relatively small '85 American history sample.

Mean differences by sex for AP history scores and college history grades

For each college with data from at least five male and five female students, score means for the AP essay, multiple-choice, and composite scores were computed separately by sex. The difference between these means (in pooled standard deviation units) was computed and weighted to give an unbiased estimate of the population value (see Hedges and Olkin, 1985 for an explanation of this procedure); this standard difference is called d in the tables. Arbitrarily, positive values of d indicate higher scores for males. Similarly, d was computed for the history grade average in each college. Again following procedures described by Hedges and Olkin, the weighted average of the ds was computed, the 95% confidence interval for this mean was determined, and the test statistic Q (which has an asymptotic chi-square distribution under the null hypothesis) was computed to test the reasonableness of the assumption that each college sample is estimating the same population effect size.

Means, standard deviations, and ds for the '84 AP American history examination and college history grades are presented in Table 7. Consistent

---

Insert Table 7 about here

---

with Mazzeo et. al. (1989), sex differences on the multiple-choice score were substantially larger than on the essay score. Course grades of males and females did not differ significantly (95% confidence interval from -.22 to .04). Note that the confidence intervals for the multiple-choice score difference and the composite score difference do not overlap with the confidence interval for the course grade difference, suggesting that these scores underestimate the ability of females to perform in college history

courses. Given that, as already noted, college history courses are primarily graded with essay tests this result is not surprising, but its practical significance should not be ignored. To put this difference in more concrete terms, suppose you had a special award that was to be given for above average performance in history. If that award were based on college history grades or AP essay scores, equal numbers of males and females would be recognized. But if it were based on AP multiple-choice scores, if half of the males received the award, then only about one-third of the females would receive it.

It might be argued that the history grades of males were artificially low because they selected more difficult history courses. However, this was not the case; analyses at the individual course level (i.e., comparing males and females enrolled in the same course) revealed the same pattern as at the history grade average level.

The large sex difference observed for the multiple-choice questions on the '84 American history examination does not suggest that males would display the same advantage on all multiple-choice tests. Scores on the verbal portion of the Scholastic Aptitude Test (SAT-V) were available for all of the students in the '84 American history sample; scores were only slightly higher for males ($\underline{d}$=.15) and the 95% confidence interval (.03 to .28) did not overlap with the confidence interval for the difference in the history multiple-choice scores. Thus, the sex difference on the AP multiple-choice history questions was significantly larger than the sex difference on SAT-V.

Tables 8 and 9 indicate that the same general pattern of sex differences

---

Insert Table 8 about here

Insert Table 9 about here

---

found in the '84 American history sample was also found in the smaller '85 American history and European history samples.

Mean differences by sex for AP English Language scores and English grades

As indicated in Table 10, there were no statistically significant sex

---
Insert Table 10 about here
---

differences on any of the AP English Language and Composition scores or on college English grades although there was a nonsignificant trend (confidence interval -.42 to .05) for the females to outperform males on the essay. This is consistent with the relatively modest question type by sex interaction noted by Mazzeo et. al. (1989).

Mean differences by sex for AP Biology scores and biology grades

Table 11 shows that, among the various indicators of performance in

---
Insert Table 11 about here
---

biology, males significantly outperformed females only on the multiple-choice AP scores, but the maximum difference in the ds (between the multiple-choice score and course grade) was only .23. For comparison, note that a difference of at least .57 between the multiple-choice d and the grade d was observed in all three history samples.

CONCLUSIONS

In terms of correlations with college freshman GPA, multiple-choice scores from the AP American History and Biology examinations appeared to be superior to the essays as predictors. When grades in history, English, and biology were the criteria, AP essay and multiple-choice scores (from the

subject-appropriate examination) were about equally good predictors, and overall sex differences in the magnitude of the correlations were trivial. But, just because two scores have approximately equal correlations with a criterion does not necessarily mean that it is a matter of indifference which score is used or how the scores are weighted to form a composite. Thus, for example, female history students would be disadvantaged by a greater reliance on multiple-choice scores both in terms of absolute numbers identified and in terms of the number identified holding ultimate success in college history courses constant. Indeed, the current results suggest that more weight should be placed on the history essays to achieve more nearly sex-fair selections for granting college credit or advanced placement. This argument assumes that college grades are themselves sex fair, but it could also be argued that college history grades are biased against males because they rely largely on essay examinations on which males do poorly relative to their performance on multiple-choice tests. However, to the extent that AP examinations should reflect what actually occurs in college classes, this argument is irrelevant.

If more weight were placed on the essay sections, efforts should be made to improve their reliabilities. The greatest improvement could be made by increasing both the number of essays written by the student and the number of readers rating each essay. Increasing either component by itself should have some positive impact on score reliability. Some improvement may also be possible by making statistical adjustments for systematic differences in the scoring standards of different raters (Braun, 1988).

Future research should include a survey of college grading practices to identify the extent to which grades in various courses are determined by in-class essays, multiple-choice tests, or other assessment techniques. In

addition, studies should be performed that could rule out construct-irrelevant influences as the determiners of sex differences on both essay and multiple-choice scores. For example, the generally neater handwriting of females could be discounted as an explanation of their relatively strong performance on essay tests if the same size sex difference were observed with typed essays. For the multiple-choice tests, a possible greater willingness of male examinees to guess when uncertain could be evaluated with a test that contained no penalty for guessing. Finally, if the sex differences on the two types of tests appear to be caused by true differences on the somewhat different constructs assessed by the two formats, then research should focus on understanding how these differences developed. Educational strategies might then be developed that could optimize performance for both sexes on essay and multiple-choice tests.

REFERENCES

Bell, R.C., & Hay, J.A. (1987). Differences and biases in English language examination formats. British Journal of Educational Psychology, 57, 212-220.

Bolger, N., & Kellaghan, T. (1990). Method of measurement and gender differences in scholastic achievement.

Braun, H.I. (1988). Understanding score reliability: Experience calibrating essay readers. Journal of Educational Statistics, 13, 1-18.

Breland, H., Camp, R., Jones, M., & Rock, D. (1987). Assessing writing skill. New York: College Entrance Examination Board.

Breland, H. & Gaynor, J. (1979). A comparison of direct and indirect assessments of writing skill. Journal of Educational Measurement, 16, 119-128.

Bridgeman, B. (1989). Comperative Validity of Multiple-Choice and Free-response items on the Advanced Placement Examination in Biology. New York: College Entrance Examination Board (CBR No. 89-2).

Dunn, O.J., & Clark, V.A. (1969). Correlation Coefficients measured on the same individuals. Journal of the American Statistical Association, 64, 366-377.

Eignor, D.R., Flesher, R.B., & McClean, D. (1984). Test Analysis: College Board Advanced Placement Examination American History. Princeton, New Jersey: Educational Testing Service. Unpublished Statistical Report No. SR-84-104.

Hedges, L.V., & Olkin, I. (1985). Statistical Methods for Meta-Analysis. New York: Academic Press Inc.

Livingston, S., Karatka, C., & Blestein, C. (1985). Test Analysis: College Board Advanced Placement Examination English Language and Composition. Princeton, New Jersey: Educational Testing Service. Unpublished Statisical Report No. SR-85-151.

Livingston S., McClean, D., & Flesher R.B. (1985). Test Analysis: College Board Advanced Placement Examination American History. Princeton, New Jersey: Educational Testing Service. Unpublished Statisical Report No. SR-85-165.

Mazzeo, J. & Flesher, R.B. (1985). Test Analysis: College Board Advanced Placement Examination European History. Princeton, New Jersey: Educational Testing Service. Unpublished Statisical Report No. SR-85-182.

Mazzeo, J. & Flesher, R.B. (1985b). Test Analysis: College Board Advanced Placement Examination Biology. Princeton, New Jersey: Educational Testing Service. Unpublished Statisical Report No. SR-85-143.

Murphy, R.J.L. (1982). Sex differences in objective test performance. British Journal of Educational Psychology, 52, 213-219.

Appendix A

Conversion from Composite Score to 1-5 Score

AP Examination	Composite Score Range	1-5 Score	% of students taking test
84 American History	0-50	1	11
	51-77	2	28
	78-94	3	27
	95-117	4	28
	118-180	5	6
American History	0-47	1	6
	48-73	2	29
	74-89	3	27
	90-111	4	26
	112-180	5	11
European History	0-60	1	10
	61-77	2	16
	78-103	3	39
	104-123	4	24
	124-180	5	11
English Language and Composition	0-49	1	15
	50-73	2	19
	74-89	3	34
	90-101	4	30
	102-150	5	4
Biology	0-35	1	11
	37-54	2	19
	55-76	3	31
	77-96	4	24
	97-150	5	14

Note: Except for the 1984 American History examination, all data is for the 1985 examinations

APPENDIX B

Detailed Tables for Correlations of AP History Examinations and Grades

Table B-1

Correlation of 1984 AP American History Scores with College History Grades

College	N		Correlation of History Grades with Essay Total			Essay1[a]		Essay2[b]		Multi. Choice			Composite		
	M	F	M	F	T	M	F	M	F	M	F	T	M	F	T
1	133	80	.47	.15	.36	.32	.14	.43	.08	.34	.09	.25	.50	.14	.37
2	50	65	.05	.42	.28	.07	.41	.02	.25	.20	.37	.27	.16	.50	.35
3	56	44	.24	.41	.31	.38	.31	.08	.33	.33	.35	.34	.34	.42	.37
4	45	35	.39	.28	.32	.36	.24	.29	.24	.12	.51	.15	.33	.45	.29
5	24	42	.36	.36	.37	.38	.38	.27	.14	.34	.30	.36	.41	.38	.42
6	42	16	.30	.15	.25	.23	.33	.29	-.05	.26	.57	.38	.33	.40	.36
7	20	31	.48	.18	.28	.50	.18	.29	.11	.32	.30	.29	.44	.31	.35
8	26	24	.10	.37	.18	.12	.40	.02	.27	.02	.08	.01	.07	.32	.13
9	30	13	.50	-.23	.36	.42	-.32	.42	.00	.37	.33	.36	.51	.09	.43
10	18	24	.30	.33	.28	.50	.14	.02	.34	.49	.60	.57	.53	.61	.58
11	25	10	-.33	.69	.01	-.21	.52	-.27	.70	.16	.56	.27	-.03	.64	.18
12	16	14	.01	.35	.18	.21	.44	-.21	.15	-.05	.51	.22	-.03	.60	.27
13	16	13	.25	.37	.30	.10	-.30	.33	.65	.54	.49	.49	.47	.53	.48
14	14	6	.38	.44	.40	.24	-.31	.26	.84	.21	.15	.35	.36	.27	.42
15	11	8	.07	.04	.10	.31	.24	-.18	-.20	.09	-.65	-.29	.09	-.40	-.11
16	8	5	.40	-.58	.14	.43	-.59	.30	-.40	.62	-.76	.21	.56	-.65	.18
17	6	9	.33	-.21	.05	.47	-.03	.08	-.29	.65	-.27	.14	.57	-.33	.13
18	6	6	-.22	-.04	-.12	.04	-.88	-.37	.58	.48	.89	.70	.21	.83	.56
Total	546	445													
Weighted Mean r			.30	.28	.29	.28	.23	.22	.21	.28	.31	.28	.36	.36	.35
95% confidence interval-low			.22	.19	.23	.20	.14	.14	.11	.20	.22	.22	.28	.27	.29
95% confidence interval-high			.38	.37	.35	.36	.33	.30	.30	.36	.40	.34	.43	.44	.40
Q (.05 level = 27.6)			24.5	14.2	9.4	13.9	24.1	24.6	17.6	10.6	29.0	22.9	17.5	23.6	14.6

[a]Essay 1 is the document essay
[b]Essay 2 is the choice essay

BEST COPY AVAILABLE

Table B-2

Correlation of 1985 AP American History Scores with College History Grades

College	N		Correlation of Course Grades with Essay Total			Essay1[a]		Essay2[b]		Multi. Choice			Composite		
	M	F	M	F	T	M	F	M	F	M	F	T	M	F	T
1	53	22	.29	.29	.28	.18	.21	.26	.29	.26	.12	.22	.34	.25	.30
2	15	6	.36	.34	.36	.21	.43	.25	.47	.19	.67	.19	.34	.62	.33
4	20	13	.26	-.16	.13	.23	-.31	.19	.01	.05	.22	.08	.20	.04	.13
5	22	38	.41	-.24	.04	.22	-.33	.43	.04	.20	.40	.30	.33	.20	.24
6	10	7	-.04	.08	.20	.22	.56	-.33	-.53	-.13	.80	.54	-.11	.70	.46
8	26	18	.47	.36	.40	.42	.43	.39	.18	.39	.52	.39	.48	.54	.46
9	9	8	.38	-.28	.16	.17	-.07	.37	-.40	-.09	-.14	-.11	.18	-.63	.04
10	13	18	.04	.12	.09	-.12	.17	.18	.02	.18	.17	.16	.14	.17	.15
11	11	6	.55	.41	.48	.52	.14	.36	.53	.12	.66	.28	.36	.61	.44
20	6	6	.80	.45	.59	.74	.31	.64	.30	-.27	.36	.16	.43	.60	.51
21	5	10	-.20	.04	-.01	-.57	-.27	-.06	.20	-.20	.40	.20	-.22	.32	.13
Total	190	152													
Weighted Mean r			.33	.05	.23	.24	.01	.29	.06	.19	.35	.24	.31	.27	.29
95% confidence interval-low			.19	-.13	.12	-.08	-.17	.14	-.12	.03	.18	.13	.16	.09	.18
95% confidence interval-high			.46	.22	.33	.38	.19	.43	.24	.33	.49	.34	.44	.42	.39
Q (.05 level = 18.3)			5.8	8.2	8.8	6.5	12.5	5.5	5.7	3.9	7.3	8.8	3.7	10.8	6.2

[a] Essay 1 is the document essay
[b] Essay 2 is the choice essay

Table B-3

Correlations of 1985 AP European History Scores with College History Grades

College	N		Correlation of History Grades with Essay			Multi. Choice			Composite		
	M	F	M	F	T	M	F	T	M	F	T
1	65	24	.36	.29	.35	.45	.35	.41	.52	.36	.47
2	12	9	-.31	-.02	-.07	.23	.17	.32	.01	.07	.18
4	19	16	-.04	-.12	-.06	-.15	.10	-.09	-.13	-.01	-.10
5	13	9	.25	-.04	.19	.21	.40	.22	.24	.20	.22
6	21	7	.33	.22	.24	.45	.75	.46	.48	.52	.42
8	14	7	.45	.66	.50	.60	.01	.33	.58	.33	.46
9	16	8	.13	.00	.09	.35	.24	.32	.29	.14	.25
Total	160	80									
Weighted Mean r			.25	.14	.22	.36	.29	.31	.39	.23	.32
95% confidence interval-low			.09	-.11	.09	.21	.04	.19	.24	-.02	.20
95% confidence interval-high			.40	.38	.35	.50	.50	.42	.52	.95	.44
Q (.05 level = 12.6)			6.21	3.6	8.17	6.9	2.96	7.51	9.2	1.9	10.3

Table 1

Correlation of Various AP Tests with College GPA

Gender	Test/n	84 American History	85 American History	European History	English Language	Biology
	Essay	.23*	.25*	.29	.25	.30*
	Multiple-Choice	.32	.31	.25	.26	.36
Combined	Composite	.33	.34	.32	.31	.36
	n -- students	3165	1054	879	821	1265
	n -- colleges	32	22	22	17	25
	Essay	.22*	.27	.32	.17	.28*
	Multiple-Choice	.33	.30	.28	.21	.37
Male	Composite	.33	.34	.36	.23	.36
	n -- students	1753	590	533	389	644
	Essay	.24*	.22*	.25	.34	.33
	Multiple-Choice	.35	.40	.29	.31	.38
Female	Composite	.36	.37	.31	.39	.39
	n -- people	1412	464	346	432	621

* Significant ($p<.05$) difference between essay and multiple choice correlations with grades

Table 2

Correlation of Various AP Tests with College Social Science Grades

Gender	Test/n	84 American History	85 American History	European History	English Language	Biology
Combined	Essay	.22*	.23*	.25	.22	.29*
	Multiple-Choice	.32	.31	.24	.23	.40
	Composite	.33	.33	.29	.28	.37
	n -- students	2292	790	599	598	922
	n -- colleges	27	19	15	15	22
Male	Essay	.21*	.25	.23	.15	.20*
	Multiple-Choice	.32	.28	.24	.17	.35
	Composite	.32	.32	.29	.20	.29
	n -- students	1221	421	354	264	444
Female	Essay	.23*	.20*	.24	.32	.38*
	Multiple-Choice	.35	.39	.30	.30	.48
	Composite	.35	.36	.32	.37	.47
	n -- people	1071	369	245	334	478

* Significant ($p<.05$) difference between essay and multiple choice correlations with grades

Table 3

Correlation of AP History Tests with College History Grades

Gender	Test/n	'84 American History	'85 American History	European History
Combined	Essay	.29	.23	.22
	Multiple-Choice	.28	.24	.31
	Composite	.35	.29	.32
	n -- students	991	342	240
	n -- colleges	18	11	7
Male	Essay	.30	.33	.25
	Multiple-Choice	.28	.19	.36
	Composite	.36	.31	.39
	n -- students	546	190	160
Female	Essay	.28	.05*	.14
	Multiple-Choice	.31	.35	.29
	Composite	.35	.27	.23
	n -- students	445	152	80

*Significant ($p<.05$) difference between essay and multiple choice correlations with grades.

Table 4

Correlation of AP History and Biology scores with SAT-V and HSGPA

	AP Scores			
	84 History		Biology	
	Essay	M-C	Essay	M-C
SAT-V	.25	.53	.28	.53
SAT-M	.10	.29	.23	.43
HSGPA	.16	.20	.15	.19

Note - History n=2837; biology n=1104

Table 5

Correlation of AP English Language Scores with College English Grades

Correrlation of English grades with

College	N M	N F	Essay M	Essay F	Essay T	Multi. Choice M	Multi. Choice F	Multi. Choice T	Composite M	Composite F	Composite T
1	29	26	.26	.23	.22	-.06	.27	.02	.12	.29	.15
2	43	62	.23	.12	.17	.21	.18	.19	.32	.17	.23
3	22	17	-.08	.38	.24	.23	-.01	.19	.11	.24	.26
4	22	23	.10	.84	.65	.37	.61	.51	.33	.83	.69
5	20	34	.08	.00	.02	.07	.07	.07	.09	.04	.05
9	8	17	.02	.39	.23	.44	.41	.36	.28	.47	.35
10	8	12	.22	-.44	.09	.18	-.20	-.02	.25	-.49	.06
11	17	22	.31	.29	.29	.11	.24	.17	.25	.35	.29
12	18	22	.43	-.02	.10	.09	.06	.08	.28	.02	.09
14	10	12	.02	.65	.31	.57	.06	.42	.37	.46	.46
15	5	5	.90	.90	.74	.36	.77	.63	.78	.85	.81
16	6	11	.47	-.01	.31	.17	.41	.29	.32	.26	.34
Total	208	273									
Weighted Mean r			.20	.24	.24	.18	.21	.20	.24	.27	.28
95% confidence interval-low			.06	.12	.16	.04	.09	.11	.10	.14	.19
95% confidence interval-high			.34	.36	.33	.32	.32	.29	.38	.38	.36
Q (.05 level = 19.7)			7.2	35.3	20.1	4.8	11.3	13.6	3.1	30.1	26.0

Table 6

Correlation of AP Biology Scores with Biology Grades

College	N		Correlation of Biology Grades with Essay			Multi. Choice			Composite		
	M	F	M	F	T	M	F	T	M	F	T
1	11	14	-.34	.63	.03	.05	.57	.33	-.22	.66	.17
2	20	27	.43	.28	.31	.55	.48	.40	.53	.47	.40
4	13	12	.35	.21	.26	.48	.25	.40	.41	.24	.33
5	10	17	-.06	.26	.19	-.01	.22	.19	-.04	.26	.20
6	14	17	.31	.74	.60	.26	.39	.29	.34	.66	.53
8	12	10	.51	.49	.51	.34	.51	.45	.45	.58	.53
11	9	27	.11	.16	.16	.32	.28	.30	.29	.25	.27
12	19	20	.42	.29	.33	.54	.21	.35	.53	.26	.37
17	9	15	.41	.47	.43	.47	.60	.54	.50	.57	.52
18	8	5	-.15	.33	-.05	-.11	-.26	-.18	-.17	.06	-.14
19	6	5	.55	-.34	-.06	.84	-.37	.00	.89	-.36	-.04
Total	131	169									
Weighted Mean $\underline{r}$			.28	.37	.30	.39	.37	.33	.37	.42	.35
95% confidence interval-low			.09	.22	.18	.21	.21	.22	.19	.27	.24
95% confidence interval-high			.45	.51	.40	.54	.50	.43	.53	.55	.45
Q (.05 level = 18.3)			7.4	9.3	11.1	7.4	5.9	6.6	10.7	7.8	9.1

Table 7

Means, Standard Deviations, and ds for 1984 AP American History Scores and College History Grades

	n		Essay					Multiple-Choice					Composite					Grade				
			M		F			M		F			M		F			M		F		
College	M	F	$\bar{X}$	SD	$\bar{X}$	SD	d	$\bar{X}$	SD	$\bar{X}$	SD	d	$\bar{X}$	SD	$\bar{X}$	SD	d	$\bar{X}$	SD	$\bar{X}$	SD	d
1	133	80	56.3	13.7	55.9	11.2	.03	58.2	10.0	53.6	9.9	.46	123.2	20.5	117.8	17.9	.27	3.24	.51	3.21	.50	.06
2	50	65	38.5	10.6	40.1	11.2	-.15	39.5	10.0	36.1	10.8	.32	85.3	17.0	83.3	18.3	.11	2.84	.67	3.15	.76	-.43
3	56	44	42.9	13.0	43.3	11.1	-.03	45.7	13.5	39.2	12.6	.49	96.4	23.4	89.9	22.1	.28	3.21	.49	3.23	.47	-.04
4	45	35	49.1	13.3	47.3	12.7	.14	50.6	10.4	44.8	10.8	.54	107.9	19.9	99.8	21.0	.39	2.94	.88	3.24	.42	-.41
5	24	42	44.9	13.4	40.9	10.2	.35	44.7	12.7	36.7	12.0	.65	97.3	23.4	84.8	20.1	.58	3.11	.57	2.70	.82	.55
6	42	16	51.5	13.3	51.6	12.5	-.01	54.0	9.5	44.1	10.8	.99	113.8	20.3	103.3	21.3	.50	3.18	.38	3.11	.55	.16
7	20	31	35.7	9.8	35.9	11.4	-.02	39.1	12.3	27.1	12.5	.95	82.1	20.0	69.5	19.7	.62	3.01	.65	2.98	.71	.04
8	26	24	46.5	11.9	42.8	14.6	.27	40.8	10.5	37.4	10.6	.32	94.7	19.7	87.3	19.6	.37	2.78	.56	3.04	.49	-.49
9	30	13	54.8	13.0	51.7	11.2	.24	51.3	12.8	50.5	12.3	.06	114.2	22.7	110.3	19.0	.18	3.28	.57	3.29	.35	-.02
10	18	24	38.5	10.1	44.0	11.5	-.49	39.2	11.8	36.2	15.8	.21	85.0	17.2	87.3	22.6	-.11	2.96	.59	2.80	.75	.23
11	25	10	34.3	8.7	37.8	12.2	-.35	31.6	13.0	35.8	12.3	-.32	72.7	19.7	80.7	24.7	-.37	2.62	.75	3.30	.48	-.96
12	16	14	49.3	9.9	43.3	14.1	.49	42.9	11.3	35.2	13.4	.61	99.8	15.9	85.6	20.3	.76	2.78	.88	2.86	.93	-.09
13	16	13	48.6	10.9	51.9	11.8	-.28	51.5	11.0	49.9	11.4	.14	108.3	19.1	109.9	19.6	-.08	3.13	.57	3.28	.37	-.46
14	14	6	46.5	8.6	37.5	8.4	1.01	46.9	10.9	31.1	12.4	1.33	101.3	16.0	75.3	19.3	1.47	2.92	.60	2.74	1.02	.23
15	11	8	42.8	10.4	45.8	8.0	-.30	47.6	8.5	43.7	9.7	.41	98.4	18.0	97.1	16.3	.07	2.83	.75	3.25	.76	-.53
16	8	5	57.8	17.0	52.2	15.5	.32	56.0	8.2	49.9	7.9	.70	122.3	22.0	110.2	23.6	.50	2.87	.56	2.97	.34	-.19
17	6	9	45.0	7.8	36.3	6.1	1.20	47.0	11.4	35.4	11.1	.97	99.8	18.6	78.8	13.6	1.26	2.85	.70	2.88	.48	-.05
18	6	6	38.5	10.5	36.0	4.2	.29	44.3	8.8	33.0	12.6	.96	90.5	10.6	75.8	14.2	1.08	2.93	.57	2.78	.66	.22
Total	546	445																				
Weighted Mean d							.05					.48					.32					-.09
95% confidence interval - low							-.08					.35					.19					-.22
95% confidence interval - high							.18					.61					.45					.04
Q (.05 level = 27.6)							18.5					18.7					21.0					23.2

Table 8

Means, Standard Deviations, and ds for 1985 AP American History Scores and College History Grades

	n		Essay					Multiple-Choice					Composite					Grade				
			M		F			M		F			M		F			M		F		
College	M	F	$\bar{X}$	SD	$\bar{X}$	SD	d	$\bar{X}$	SD	$\bar{X}$	SD	d	$\bar{X}$	SD	$\bar{X}$	SD	d	$\bar{X}$	SD	$\bar{X}$	SD	d
1	53	22	54.0	11.6	50.0	17.6	.29	58.8	8.7	53.6	9.6	.57	119.1	16.9	109.5	25.1	.48	3.26	.42	3.23	.39	.07
2	15	6	44.6	14.1	46.0	11.3	-.10	45.6	15.5	33.9	14.4	.74	95.5	24.4	84.3	23.2	.45	3.26	.75	3.52	.45	-.36
4	20	13	46.4	12.0	47.5	13.0	-.09	46.9	10.3	40.7	12.9	.53	98.6	18.3	93.2	23.5	.26	3.05	.50	3.11	.31	-.13
5	22	38	39.4	9.8	38.0	8.4	.15	38.2	13.1	32.0	12.4	.48	82.3	21.1	74.2	17.1	.43	2.59	.91	2.68	.79	-.11
6	10	7	50.7	9.8	41.1	4.1	1.13	52.5	8.1	42.7	11.6	.96	109.0	14.2	80.9	14.8	1.32	3.12	.46	2.63	.81	.75
8	25	18	41.9	9.9	40.7	14.3	.10	42.7	11.6	38.6	11.5	.35	89.7	19.1	84.0	21.6	.28	2.73	.56	2.94	.50	-.38
9	9	8	53.0	10.7	56.6	9.8	-.33	49.7	11.1	50.3	8.6	-.06	108.3	16.2	112.6	6.5	-.32	3.28	.56	3.32	.35	-.08
10	13	18	44.1	12.9	45.0	15.4	-.06	44.7	14.3	37.8	10.0	.56	94.0	24.1	87.5	20.8	.28	2.93	.55	2.95	.57	-.03
11	11	6	37.4	11.0	36.0	11.4	.12	36.6	11.0	40.3	7.5	-.35	78.5	20.5	81.2	16.6	-.13	2.77	.75	3.00	.89	-.27
20	6	6	41.5	16.5	43.0	11.6	-.10	53.3	11.8	34.9	7.5	1.72	100.7	22.4	82.3	13.7	.91	2.94	.43	2.75	.48	.38
21	5	10	36.0	8.2	39.9	10.5	-.37	31.8	8.5	37.0	8.7	-.57	72.0	15.5	81.5	12.7	-.66	2.07	1.45	2.44	.97	-.31
total	190	184																				
Weighted Mean $\bar{d}$							.11					.46					.34					-.06
95% confidence interval - low							-.11					.23					.12					-.28
95% confidence interval - high							.34					.69					.57					.17
Q (.05 level = 18.3)							.68					12.1					10.6					4.6

Table 9

Means, Standard Deviations, and ds for European History Scores and College History Grades

	n		Essay					Multiple-Choice					Composite					Grade				
			M		F			M		F			M		F			M		F		
College	M	F	$\bar{X}$	SD	$\bar{X}$	SD	d	$\bar{X}$	SD	$\bar{X}$	SD	d	$\bar{X}$	SD	$\bar{X}$	SD	d	$\bar{X}$	SD	$\bar{X}$	SD	d
1	65	24	61.6	11.7	65.5	11.8	-.33	64.3	14.5	59.1	14.5	.36	125.9	20.9	124.6	23.9	.06	3.26	.49	3.38	.42	-.15
2	12	9	55.8	9.2	55.3	15.2	.04	63.1	12.9	43.9	12.4	1.45	118.8	17.4	99.2	27.0	.86	3.26	.51	2.57	1.56	.61
4	19	16	54.8	9.3	56.8	8.1	-.22	56.7	11.2	52.2	8.5	.44	111.5	15.3	109.0	13.1	.17	3.02	.68	3.05	.37	-.05
5	13	9	49.8	20.2	50.7	12.6	-.05	46.1	19.5	41.3	13.2	.27	95.9	37.9	92.0	23.6	.11	2.51	1.01	2.79	.64	-.31
6	21	7	64.1	12.1	60.0	13.3	.32	65.5	11.2	63.7	9.3	.16	129.6	18.5	123.7	19.1	.31	3.24	.49	3.57	.43	-.67
8	14	7	46.9	12.8	51.0	7.7	-.34	49.2	15.0	45.3	11.2	.27	92.3	23.0	96.3	16.0	-.18	2.79	.63	3.27	.66	-.72
9	16	8	60.2	12.9	63.4	13.5	-.24	60.6	15.1	54.0	14.6	.43	120.8	23.8	117.4	24.1	.14	3.39	.46	3.35	.34	.09
Total	160	80																				
Weighted Mean d							-.17					.43					.17					-.15
95% confidence interval - low							-.44					.15					-.11					-.42
95% condfidence interval - high							.10					.70					.44					.12
Q (.05 level = 12.6)							2.19					5.00					3.14					6.18

Table 10

Means, Standard Deviations, and ds for AP English Scores and English Grades

	n		Essay					Multiple-Choice					Composite					Grade				
			M		F			M		F			M		F			M		F		
College	M	F	$\bar{X}$	SD	$\bar{X}$	SD	d	$\bar{X}$	SD	$\bar{X}$	SD	d	$\bar{X}$	SD	$\bar{X}$	SD	d	$\bar{X}$	SD	$\bar{X}$	SD	d
1	29	26	59.0	9.9	57.2	9.3	.18	47.2	8.0	45.9	6.8	.17	106.1	16.8	103.2	14.2	.18	3.19	.48	3.37	.28	-.46
2	43	62	46.4	9.2	47.6	8.9	-.13	33.3	8.3	29.0	8.8	.50	79.8	12.5	76.6	15.2	.22	3.00	.57	2.99	.52	.02
3	22	17	45.6	7.1	52.7	10.7	-.79	33.3	8.8	36.6	9.7	-.35	78.9	12.5	89.4	17.2	-.70	2.81	.61	3.24	.42	-.79
4	22	23	54.1	6.7	53.2	14.4	.09	42.1	7.5	42.4	8.5	-.04	96.3	10.8	95.6	20.7	.04	3.18	.53	2.98	.77	.30
5	20	34	43.7	10.9	47.6	8.4	-.47	33.0	8.5	34.1	7.8	-.13	76.0	15.8	81.8	13.0	-.41	2.74	.79	2.64	.73	.13
9	8	17	50.5	8.2	56.1	9.3	-.60	36.9	7.3	41.1	10.4	-.46	87.1	13.3	97.1	16.6	-.62	3.34	.31	3.18	.42	.40
10	8	12	43.8	13.2	46.1	6.5	-.23	32.6	5.0	31.6	9.0	.12	76.5	14.7	77.8	9.4	-.11	3.06	.82	3.18	.30	-.20
11	17	22	49.0	11.9	48.2	7.6	.08	34.0	11.9	34.3	9.4	-.03	83.0	20.4	82.6	12.5	.02	3.25	.65	3.34	.63	-.14
12	18	32	43.9	9.9	47.2	7.9	-.38	34.4	12.2	32.9	10.7	.13	78.6	18.0	80.2	16.2	-.09	3.36	.84	2.89	.99	.49
14	10	12	46.4	8.0	49.3	9.4	-.32	31.0	9.2	39.5	9.4	-.88	77.5	14.8	88.9	14.4	-.75	2.37	.95	2.80	.53	-.55
15	5	5	48.2	7.6	49.2	7.3	-.12	38.8	9.8	31.2	6.6	.82	86.4	13.7	80.4	13.7	.39	3.40	.55	2.77	.72	.89
16	6	11	37.8	12.7	50.6	8.0	-1.23	35.0	14.9	35.5	9.1	-.04	72.8	26.4	86.1	14.1	-.66	3.25	.46	3.46	.41	-.47
Total	208	273																				
Weighted Mean d							-.23					.05					-.11					-.04
95% confidence interval - low							-.42					-.12					-.29					-.23
95% condfidence interval - high							.05					.24					.07					.14
Q (.05 level = 19.7)							12.5					14.4					13.8					17.3

BEST COPY AVAILABLE

TABLE 11

Means, Standard Deviations, and ds for AP Biology Scores and Biology Grades

			Essay					Multiple-Choice					Composite					Grade				
	n		M		F			M		F			M		F			M		F		
College	M	F	$\bar{X}$	SD	$\bar{X}$	SD	d	$\bar{X}$	SD	$\bar{X}$	SD	d	$\bar{X}$	SD	$\bar{X}$	SD	d	$\bar{X}$	SD	$\bar{X}$	SD	d
1	11	14	46.0	16.6	42.5	10.1	.25	56.0	9.0	59.1	10.8	.18	102.2	23.6	96.6	18.9	.26	3.34	.72	3.31	.55	.05
2	20	27	29.9	14.0	28.3	10.3	.13	41.9	11.6	41.7	8.3	.02	71.7	23.1	69.8	14.8	.10	2.74	.83	2.60	.63	.19
4	13	12	36.3	16.6	36.5	12.4	-.01	49.3	13.2	46.0	7.1	.30	85.5	28.8	82.7	17.8	.11	2.92	.72	2.29	.80	.80
5	10	17	22.9	9.0	17.5	11.1	.50	38.2	10.2	30.9	11.2	.65	61.3	18.3	48.1	19.8	.66	2.53	1.24	2.01	.96	.47
6	14	17	45.1	6.2	42.8	11.3	.24	58.3	5.7	53.7	9.5	.56	103.3	9.0	96.5	18.3	.45	2.71	.66	2.91	.87	-.25
8	12	10	33.9	12.5	30.2	13.8	.27	48.2	9.3	44.2	11.3	.38	82.0	20.7	74.2	21.3	.36	2.83	.64	2.57	.72	.37
11	9	27	23.4	9.7	25.1	10.8	-.16	35.8	8.4	38.0	10.3	-.22	59.2	13.3	63.1	18.6	-.22	1.33	1.00	2.37	.85	-1.15
12	19	20	22.1	10.0	24.9	12.4	-.24	40.9	10.2	37.5	10.7	.32	63.1	18.0	62.4	21.6	.03	2.97	.92	2.91	1.07	.06
17	9	15	16.6	8.2	20.5	14.2	-.30	30.6	12.5	33.7	14.4	-.22	47.0	19.0	54.2	26.9	-.29	2.87	1.31	2.85	1.11	.02
18	8	5	22.8	10.1	25.6	11.2	-.25	37.6	7.3	31.4	9.2	.72	60.4	15.0	51.2	16.9	.19	2.34	1.33	2.80	.27	-.40
19	6	5	37.7	10.6	28.0	7.9	.93	49.5	6.7	43.8	5.0	.87	87.3	13.0	71.8	12.5	1.11	2.53	.27	2.63	.80	-.16
total	131	169																				
Weighted Mean g							.08					.26					.19					.03
95% confidence interval - low							-.15					.02					-.05					-.20
95% condfidence interval - high							.32					.49					.42					.27
Q (.05 level = 18.3)							2.62					2.25					2.21					4.58